Integration of Metadata in BWmeta-2.0.0 Format

Katarzyna Zamlynska†, Jakub Jurkiewicz‡ and Lukasz Bolikowski

University of Warsaw, Interdisciplinary Centre for Mathematical and Computational Modelling, Warsaw, Poland
† kzaml(at)icm.edu.pl
‡ kura(at)icm.edu.pl

Abstract: *This article describes the BWmeta-2.0.0 format in terms of its use in an integrated digital library.*

Keywords: *Mathematical digital libraries, Metadata indexing format, YADDA*

I. INTRODUCTION

YADDA [1] is a document repository framework for digital publication of scientific works. It is built on an open architecture and therefore is flexible and easily adapts to the particular needs of publishers. It can aggregate contents from a diverse range of sources of varying data type, not only publications, but also programs, experimental results, movies etc. At present, in YADDA repository [2], 5 different collections are available. Those collections cover a wide range of disciplines from humanities to science to mathematics, containing both books and journals, with different access regimes. The variety of content to be held in the YADDA repository is on the one hand very attractive for users, while on the other hand challenging for programmers [3]. Proper description of the content, clear and detailed, is not an easy task. To obtain the required level of flexibility the new metadata format, called BWmeta, was implemented in the YADDA platform. The latest release of BWmeta is BWmeta-2.0.0, which has a range of new features, which further improve our digital library.

II. HIERARCHY

The Polish Digital Mathematical Library (PL DML) [4] contains both books and journals. In one of the most popular metadata format for scientific publications tag libraries for books and journals are separated. In BWmeta, description of an object does not depend on its type. Chapters of books and journal articles are all described with the same set of elements. The hierarchy of documents is not strictly defined. It can be:

- publisher – book series – volume – book – chapter
- publisher – journal – volume – issue – article
- series – conference – article

but also any other suitable to the situation and particular needs of a contributor, like: institution – Ph.D. thesis. Another flexibility is the ability to add detailed metadata to every level of hierarchy. It is very useful in journals like Banach Center Publications from the PL DML collection, in which every volume has its own title and editors.

```
<structure hierarchy="BWmeta1.hierarchy-
class.hierarchy_Journal">
<level="BWmeta1.level.hierarchy_Journal_Article"
position="169-178"/>
<ancestor
level="BWmeta1.level.hierarchy_Journal_Journal">
<name lang="eng">Banach Center
Publication</name> <contributor institution="false"
role="editor">
<name>Andrzej Hulanicki</name>
</contributor> </ancestor> <ancestor
level="BWmeta1.level.hierarchy_Journal_Volume">
<name>40</name> <name lang="eng"
type="subtitle">Mathematics of
Gravitation</name> </ancestor> <ancestor
level="BWmeta1.level.hierarchy_Journal_Issue">
<name>1</name> <name lang="eng"
type="subtitle">Lorentzian Geometry and Einstein
Equations</name
<contributor institution="false" role="editor">
<name>Piotr Chru´sciel</name> </contributor>
</structure> ... <name lang="eng">A stable class of
spacetimes with naked singularities</name>
<contributor institution="false" role="author">
<name type="surname">Kriele</name>
</contributor>
```

This feature is also helpful in multi-volume books like Ouevres of Stefan Banach, in which every volume has its own editors, title and every chapter can have different authors and date of publications.

Each object can belong to more than one hierarchy class. To describe the place in hierarchy one can use the reference to higher object. References can also be done in other elements like contributor or affiliation. It is enough to add the attribute id, which allows the element content to be referenced. For example:

```
<affiliation id="ICM_UW"> <text>Interdisciplinary
Centre for Mathematical and Computational Modelling,
University of Warsaw</text> </affiliation>
```

The reference to the ”ICM UW” can be done later by the element:

```
<affiliation-ref ref="ICM_UW"/>
```

In BWmeta-2.0.0 an identifier is not an XML reference. The definition of the reference element does not

have to be declared in the same XML file. Moreover, during the import to the system the identifiers don't have to be available in the import package, like it was required in BWmeta-1.2.0.

III. DIVERSITY OF METADATA

Once the metadata scheme is defined, adding new tags to it creates problems. Instead of trying to predict all possible types of data, we decided to make the tag system as easy as possible. The detailed description of elements can be done by adding parameters. In the PL DML library, few types of description exist. All of them are marked by <description>. To distinguish between an abstract written by author or a comment add by the editor, it is enough to add the attribute type:

```
<description lang="eng" type="abstract">
```

where type can be: abstract, summary, comment, note. If it is needed a new type can be added, for example the table of content in a book, without redefining metadata or changing software presentation. In BWmeta-2.0.0 elements are not strictly defined, they can describe different types of data. For example: the original title and it's translation are described by the same element <name>, and distinctions between them can be made by attribute type:

```
<name lang="eng">Quantum Mechanics</name>
<name lang="pl" type="translated">Mechanika
Kwantowa</name>
```

The attribute type doesn't have to be included. Most of the XML elements can be expanded with information which doesn't suit the schema. It can be done with the help of the element attribute. For example with citations, instead of creating a number of new tags, adding to the element <attribute> parameter key, we are able to keep in the same place the unstructured an structured citation forms.

```
<relation type="reference">
<attribute key="text">
<value>1. Winfree, A. T. The Geometry of Biological
Time
(Springer, New York, 1980)."</value>
</attribute>
<attribute key="parsed">
<attribute
key="type"><value>book</value></attribute>
<attribute
key="position"><value>1</value></attribute>
<attribute key="authors">
<attribute key="author">
<value>A. T. Winfree</value>
<attribute key="forenames"><value>A.
T.</value></attribute>
<attribute
key="surname"><value>Winfree</value></attribute>
</attribute>
</attribute>
<attribute key="title">
<value>The Geometry of Biological Time</value>
</attribute> <attribute
key="publisher"><value>Springer</value></attribute
>
<attribute key="city"><value>New
York</value></attribute>
<attribute
key="year"><value>1980</value></attribute>
</attribute> </relation>
```

IV. EMEBED TEXT FORMULA

In BWmeta-2.0.0 we decided that we don't want to limit text fields to any specific XML. We would like to allow adding to text: XHTML, MathML and probably some formats we don't know about. We use for text fields XSD type: xsd:any with namespace="##other". This allows us to easily add both XHTML and MathML. Below is example of HTML:

```
<name lang="eng">
Bcl-2-dependent modulation
of Ca<sup
xmlns="http://www.w3.org/1999/xhtml">2+</sup>
homeostasis and store-operated channels in
prostate cancer cells </name>
```

Thanks to this solution we get a nice view of calcium ion: Ca^{2+} . Using same construction we input MathML:

```
<tags type="keyword" lang="eng">
   <tag>Euler</tag>
   <tag>Adding</tag>
   <tag>
                        <math
  xmlns="http://www.w3.org/1998/Math/MathML">
        <mi>summand1</mi>
        <mo>+</mo>
        <mi>summand2</mi>
        <mo>=</mo>
        <mi>sum</mi>
 </mat
 h>
 </tag
 >
 </tags
 >
```

What is important here is that we don't put MathML to CDATA section. This means that during the validation process of BWmeta XML documents, MathML, which is inside, is also validated.

V. CONCLUSIONS

The flexibility of BWmeta-2.0.0 makes it a more appropriate format for integrated libraries. It allows for easy addition of new attributes, descriptions, and names of elements, without the need to change the format of the metadata definition, presentation software and storage software. This flexibility is extremely important

during migration processes from one library system to another without loosing the elements which are not present in the new system.

ACKNOWLEDGEMENTS

This work is partially supported by the National Centre for Research and Development (NCBiR) under Grant No. SP/I/1/77065/10 by the Strategic scientific research and experimental development program: SYNAT - "Interdisciplinary System for Interactive Scientific and Scientific-Technical Information"

REFERENCES

YADDA
http://yaddainfo.icm.edu.pl/?page%20id=46

The Virtual Library of Science http://yadda.icm.edu.pl

Katarzyna Zam lynska, Lukasz Bolikowski, Tomasz Rosiek Migration of the Mathematical Collection of Polish Virtual Library of Science to the YADDA platform. DML 2008 Towards Digital Mathematics Library, Birmingham, UK, July 27th, 2008 Proceedings , Masaryk University, r. 2008

The Polish Digital Mathematical Library
http://pldml.icm.edu.pl/mathbwn/search/article.action